**Conversational Networks for Automatic Online Moderation**

**Abstract:**

Moderation of user-generated content in an online communityisachallengethathasgreatsocio-economicalramiﬁcations.However, the costs incurred by delegating this work to human agents are high. For this reason, an automatic system able to detect abuse in usergenerated content is of great interest. There are a number of ways to tackle this problem, but the most commonly seen in practice are word ﬁltering or regular expression matching. The main limitations are their vulnerability to intentional obfuscation on the part of the users, and their context-insensitive nature. Moreover, they are language-dependent and may require appropriate corpora for training. In this paper, we propose a system for automatic abuse detection that completely disregards message content. We ﬁrst extract a conversational network from raw chat logs and characterize it through topological measures. We then use these as features to train a classiﬁer on our abuse detection task. We thoroughly assess our system on a dataset of user comments originating from a French Massively Multiplayer Online Game. We identify the most appropriate network extraction parameters and discuss the discriminative power of our features, relatively to their topological and temporal nature. Our method reaches an F-measure of 83.89 when using the full feature set, improving on existing approaches. With a selection of the most discriminative features, we dramatically cut computing time while retaining most of the performance (82.65).

**SYSTEM REQUIREMENTS:**

**HARDWARE REQUIREMENTS:**

* System : Pentium Dual Core.
* Hard Disk : 120 GB.
* Monitor : 15’’ LED
* Input Devices : Keyboard, Mouse
* Ram : 1 GB

**SOFTWARE REQUIREMENTS:**

* Operating system : Windows XP/UBUNTU.
* Implementation : NS2
* NS2 Version : 2.28
* Front End : OTCL (Object Oriented Tool Command  Language)
* Tool : Cygwin (To simulate in Windows OS)